Theory for the Emergence of Modularity in Complex Systems

Jeong-Man Park and Michael W. Deem

Department of Physics & Astronomy

Rice University, Houston, TX 77005–1892, USA

Department of Physics, The Catholic University of Korea, Bucheon 420-743, Korea

Abstract

Biological systems are modular, and this modularity evolves over time and in different environments. A number of observations have been made of increased modularity in biological systems under increased environmental pressure. We here develop a theory for the dynamics of modularity in these systems. We find a principle of least action for the evolved modularity at long times. In addition, we find a fluctuation dissipation relation for the rate of change of modularity at short times.

PACS numbers: 87.10.-e, 87.15.A-, 87.23.Kg

Biological systems have long been recognized to be modular. In 1942 Waddington presented his now classic description of a canalized landscape for development, in which minor perturbations do not disrupt the function of developmental modules [1]. In 1961 H. A. Simon described how biological systems are more efficiently evolved and are more stable if they are modular [2]. A seminal paper by Hartwell *et al.* firmly established the concept of modularity in cell biology [3]. Systems biology has since provided a wealth of examples of modular cellular circuits, including metabolic circuits [4] and modules on different scales, *i.e.* modules of modules [5]. Protein-Protein interaction networks have been observed to be modular [6–8]. Ecological food webs have been found to be modular [9]. The gene regulatory network of the developmental pathway exhibits modules [10, 11], and the developmental pathway is modular [12]. Modules have even been found in physiology, specifically in spatial correlations of brain activity [13, 14].

The modularity of a complex biological system can not only be quantified, but also change over time. There are a number of demonstrations of the evolution of modularity in biological systems. For example, the modularity of the protein-protein interaction network significantly increases when yeast are exposed to heat shock [15], and the modularity of the protein-protein networks in both yeast and E. coli appears to have increased over evolutionary time [16]. Additionally, food webs in low-energy, stressful environments are more modular than those in plentiful environments [17], arid ecologies are more modular during droughts [18], and foraging of sea otters is more modular when food is limiting [19]. The modularity of social networks changes over time: stock brokers instant messaging networks are more modular under stressful market conditions [20], socio-economic community overlap decreases with increasing stress, and criminal networks are more modular under increased police pressure [21]. Modularity of financial networks changes over time: the modularity of the world trade network has decreased over the last 40 years, leading to increased susceptibility to recessionary shocks [22], and increased modularity has been suggested as a way to increase the robustness and adaptability of the banking system [23].

In an effort to explain some of these observations, we here present a theory to describe dynamics of modularity. This analytical theory complements numerical models that have investigated the dynamics of modularity [24–27]. We assume that modularity can be quantified in the system under study. We further consider that modularity is a good order parameter to describe the state of the system. That is, we project the dynamics onto the

slow mode of modularity, M. We then consider the equations of motion for the modularity of the system. In particular, we consider an ensemble of systems, each with different values of the modularity, and each evolving. The evolutionary dynamics of this system is fully specified by the rate at which systems reproduce, f, termed "fitness," and the rate at which changes of modularity arise, μ. Since the state of the system is specified by the slow variable M, the fitness is a function of the modularity, $f = f(M)$. The $f(M)$ function is from a detailed calculation, numerical simulation, or experimental observation.

We further consider that there is a pressure on this ensemble of systems to have an efficient response function. A canonical form of this pressure is a changing environment. As the environment changes, the favorable niches for the system change, and the system must adapt to the changing landscape. The more rapidly the environment changes or the more dramatically the environment changes, the more pressure there is on the system to be adaptable. As noted above, it has been widely observed that systems under pressure tend to become more modular. If we denote the rate of change of environment as $1/T$ and the magnitude of the change as p, the mean fitness of the population of systems will depend on these parameters, as well as the modularity: $f = f_{p,T}(M)$. Evolution of modularity depends on how the response function of the system varies with these parameters. Since systems under stress tend to become more modular, it is reasonable to assume that the population average fitness for a modular system is greater than that for a non-modular system, at least for small T or large p where stress is large. This behavior has been observed in a model system evolving in a changing environment, when horizontal gene transfer is included [24]. We also note that this canonical behavior has also been observed in energy relaxation dynamics of spin glass models of different sizes [28]. Glassy evolutionary dynamics has been noted a number of times [29, 30]. Conversely, at long times, the non-modular system should have a higher fitness, because modularity is a constraint on the optima that can be achieved. This is the reason for the crossing of the solid and dashed curves in Fig. 1. We here take this function $f = f_{p,T}(M)$ as input. We assume only that this function for large M and small M looks like the dashed and solid curves in Fig. 1. Putting these points together, we expect the emergence of modularity at small p or large T. We seek a theory to quantitatively describe this emergence. For very slow rates of environmental change, the system can relax to the nearly optimal configuration, which is unlikely to be modular, as modularity is a restriction on the system. Thus, we expect the population average fitness for the non-modular system

to be greater than that for a modular system for large T or small p.

To proceed further, we define the "connection matrix" for our system. The connection matrix gives the links between the nodes of the network. For example, in the protein-protein interaction network, the nodes are the proteins and the links tell one whether protein i interacts with protein j. The connection matrix Δ_{ij} is a binary matrix which denotes whether nodes i and j interact ($\Delta_{ij} = 1$) or not ($\Delta_{ij} = 0$). In the criminal network, the connection matrix denotes whether criminal i interacts with criminal j. The fitness function that underlies the detailed dynamics which define $f_{p,T}(M)$ may well have non-trivial couplings between nodes [24], and the connection matrix is the projection of the non-zero couplings. We assume that each node is connected to C other nodes on average. The number of nodes is denoted by N. Rearrangement of the ones within this matrix changes the modularity of the matrix. For simplicity, we assume that the modules which form are of size L. Thus a modular system will have excess of connections along the $L \times L$ block diagonals of the connection matrix. In other words, the probability of a connection is C_0/N outside the block diagonals when $\lfloor i/L \rfloor \neq \lfloor j/L \rfloor$ and C_1/N inside the block diagonals when $\lfloor i/L \rfloor = \lfloor j/L \rfloor$, with $C = C_0 + (C_1 - C_0)L/N$. Modularity is defined by the excess of connections in the block diagonals, over that observed outside the block diagonals: $M = (C_1 - C_0)L/(NC)$.

If our population of systems is large, i.e. we have a large biological population size, the probability distribution to have a matrix with modularity m obeys

$$\begin{aligned}\frac{dP_m(t')}{dt'} &= [f_{p,T}(m) - \langle f \rangle]P_m(t') + \mu C \frac{L}{N}\left[(1-m)\left(1-\frac{L}{N}\right) + \frac{1}{N}\right] P_{m-1/(N-L)}(t') \\ &+\mu C\left(1-\frac{L}{N}\right)\left[m + (1-m)\frac{L}{N} + \frac{1}{N}\right] P_{m+1/(N-L)}(t') \\ &-\mu C\left(1-\frac{L}{N}\right)\left(m + 2(1-m)\frac{L}{N}\right) P_m(t') \end{aligned} \quad (1)$$

where m takes values $-L/(N-L), (-L+1)/(N-L), (-L+2)/(N-L), \ldots, 1$. The average fitness is given by $\langle f(t') \rangle = \sum_m f_{p,T}(m) P_m(t')$. The average modularity as a function of time is given by $M(t') = \sum_m m P_m(t')$. Multiplying this equation by m and summing, we find that the rate of change of modularity satisfies

$$M' = \langle m f(m) \rangle - M\langle f \rangle - \mu C M/N \quad (2)$$

Here M is the average modularity of the population, and m is the modularity for any particular matrix in the population, *i.e.* $M = \langle m \rangle$. Biologists would term this equation a

continuous-time Price equation [31], and we will show below that this equation implies a type of useful fluctuation-dissipation theorem.

For large values of N, for which the changes in M are nearly continuous, the average fitness implied by Eq. (1) at long times may be determined by techniques borrowed from quantum field theory [32, 33]. The average modularity follows a dynamical trajectory away from an initial state to a final steady state value. The remarkable result from this derivation is that the modularity which emerges at long times obeys

$$f_{\rm pop} = \max_{\xi} \left\{ f_{p,T}(\xi) - \mu C[(N-L)L/N^2][2 + (N/L-2)\xi - 2\sqrt{(1-\xi)(1+(N/L-1)\xi)}] \right\} \quad (3)$$

with modularity determined by the solution of the implicit equation

$$f_{p,T}(M) = f_{\rm pop} \quad (4)$$

Here $f_{\rm pop}$ is the mean population fitness divided by N. Thus, a principle of least action gives the evolved modularity.

While Eq. (3) is a general result, we can proceed further in the limit that evolved modularities are small. Expanding for small M, we find

$$\begin{aligned} \xi_{\rm max} &= \frac{2L\left[df_{p,T}/dM|_{M=0}\right]}{\mu C(N-L) - 2L\left[d^2 f_{p,T}/dM^2|_{M=0}\right]} \\ f_{\rm pop} &= \frac{L\left[df_{p,T}/dM|_{M=0}\right]^2}{\mu C(N-L) - 2L\left[d^2 f_{p,T}/dM^2|_{M=0}\right]} + f_{p,T}(0) \\ M &= \frac{L\left[df_{p,T}/dM|_{M=0}\right]}{\mu C(N-L) - 2L\left[d^2 f_{p,T}/dM^2|_{M=0}\right]} \end{aligned} \quad (5)$$

Thus, as long as a modular system has a higher fitness, modularity will spontaneously emerge for large enough system sizes, N.

We now derive a relationship between the rate of growth of modularity and the environmental pressure. Let us say that the fitness for small values of modularity can be expressed as $f(m)/N = f_0 + m\Delta f$. Equation (2) becomes $M' = \sigma_M^2 \Delta f - \mu CM/\mathrm{N}$, where $\sigma_M^2 = \langle m^2 \rangle - M^2$. For small L/N, this equation combined with Eq. 5 implies that at steady state $\sigma_{M_\infty}^2 = L/N^2$. Let us investigate the growth of modularity from an initially non-modular state. The value of Δf depends on p. If $p = 0$, the environment is not changing, and the system will stay in the $M = 0$ state with $\Delta f = 0$. If $p = 1$ then $\Delta f = f_1 - f_0 \approx f_1$ because only the modular system can evolve significantly during the time T on the completely randomized, new landscape. Making a linear interpolation, we find $\Delta f \approx p f_1(t_c) = p f_0(t_c)$,

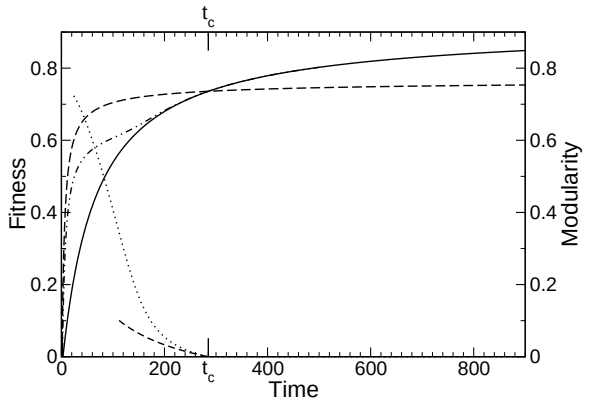

FIG. 1: Shown is the fitness of an evolving system. The fitness of the non-modular (f_0, solid), modular (f_M, dot-dashed), and block-diagonal (f_1, dashed) systems are shown. The modularity calculated from Eq. 3 is shown (dotted). Also shown is the result for small M, Eq. 5, to first order in L/N (short dashed). In this example $N = 120$, $L = 10$, $\mu = 0.01$, and $C = 12$. In this case $t_c \approx 285$.

where t_c is the time at which the non-modular and fully modular curves cross. We thus find $M' \approx \sigma_M^2 p f_0(t_c)$. Reverting back to real time, *i.e.* $t = Tt'$ because there is "fast" dynamics that occurs between each environmental change of duration T, we find

$$p_{\rm E} = \frac{1}{R}\frac{dM}{dt} \tag{6}$$

where $p_{\rm E} = p/T$ is the environmental pressure, and $R = \sigma_M^2 f_0(t_c)$.

This Eq. (6) follows from the principle of least action (3), the dynamic generalization of it, Eq. (2), and the response function of the modular system being greater than that of the non-modular system at short times. Equation (6) may be interpreted as a Taylor series expansion of dM/dt in allowed combinations of p and $1/T$. Alternatively, Eq. (6) may be interpreted as the linear response of the modularity to the environmental pressure. The coefficient R is a measure of ruggedness, since R is proportional to the variance of the modularity, which is expected to be related to the ruggedness of the landscape. The coefficient R is also expected to be related to replicate variability in experiments [34].

What does this theory mean? Equation (6) says that an increase of environmental pressure should lead to the evolution of systems with increased modularity. A study of 117 species of bacteria showed that the modularity of the bacteria's metabolic networks increased monotonically with variability of the environment in which the bacteria lived [35].

Metabolic networks of pathogens alternating between hosts were found to be more modular than those of single-host pathogens [36]. A number of other examples were mentioned in the introduction.

The present theory should allow the analysis of complex, evolving systems to go beyond a demonstration of the existence of modularity to a quantitative analysis of the dynamics of modularity.

[1] C. H. Waddington, Nature **150**, 563 (1942).

[2] H. A. Simon, Proc. Amer. Phil. Soc. **106**, 467 (1962).

[3] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, Nature **402**, C47 (1999).

[4] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002), URL http://www.sciencemag.org/cgi/content/abstract/297/5586/1551.

[5] M. R. da Silva, H. Ma, and A.-P. Zeng, Pr. Inst. Electr. Elect. **96**, 1411 (2008), URL http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4567408.

[6] V. Spirin and L. A. Mirny, Proc. Natl. Acad. Sci. USA **100**, 12123 (2003), URL http://www.pnas.org/cgi/content/abstract/100/21/12123.

[7] A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, et al., Nature **440**, 631 (2006), URL http://www.ncbi.nlm.nih.gov/pubmed/16429126.

[8] C. von Mering, E. M. Zdobnov, S. Tsoka, F. D. Ciccarelli, J. B. Pereira-Leal, C. A. Ouzounis, and P. Bork, Proc. Natl. Acad. Sci. USA **100**, 15428 (2003), URL http://www.pnas.org/cgi/content/abstract/100/26/15428.

[9] A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, Nature **426**, 282 (2003), URL http://www.ncbi.nlm.nih.gov/pubmed/14628050.

[10] E. C. Raff and R. A. Raff, Evol. Dev. **2**, 235 (2000), URL http://www3.interscience.wiley.com/journal/119052048/abstract.

[11] G. P. Wagner, Integr. Comp. Biol. **36**, 36 (1996), URL http://icb.oxfordjournals.org/cgi/content/abstract/36/1/36.

[12] C. P. Klingenberg, Annu. Rev. Ecol. Evol. S. **39**, 115 (2008), URL http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=35967311.

[13] D. Meunier, S. Achard, A. Morcom, and E. Bullmore, NeuroImage **44**, 715 (2009), URL http://www.ncbi.nlm.nih.gov/pubmed/19027073.

[14] M. Chavez, M. Valencia, V. Navarro, V. Latora, and J. Martinerie, Phys. Rev. Lett. **104**, 118701 (2010), URL http://prl.aps.org/abstract/PRL/v104/i11/e118701.

[15] A. Mihalik and P. Csermely, PLoS Comput. Biol. **7**, e1002187 (2011).

[16] J. He, J. Sun, and M. W. Deem, Phys. Rev. E **79**, 031907 (2009).

[17] D. M. Lorenz, A. Jeng, and M. W. Deem, Phys. Life Rev. **8**, 129 (2011).

[18] M. Rietkerk, S. C. Dekker, P. C. de Ruiter, and J. van de Koppel, Science **305**, 1926 (2004).

[19] M. T. Tinker, G. Bentall, and J. A. Estes, Proc. Natl. Acad. Sci. USA **105**, 560 (2008).

[20] S. Saavedra, K. Hagerty, and B. Uzzi, Proc. Natl. Acad. Sci. USA **108**, 5296 (2011).

[21] M. Kenney, in *Networked politics: Agency, power, and governance*, edited by M. Kahler (Cornell University Press, 2009), pp. 79–102.

[22] J. He and M. W. Deem, Phys. Rev. Lett. **105**, 198701 (2010), URL http://prl.aps.org/abstract/PRL/v105/i19/e198701.

[23] A. G. Haldane and R. M. May, Nature **469**, 351 (2011).

[24] J. Sun and M. W. Deem, Phys. Rev. Lett. **99**, 228107 (2007).

[25] H. Lipson, J. B. Pollack, and N. P. Suh, Evolution **56**, 1549 (2002).

[26] N. Kashtan and U. Alon, Proc. Natl. Acad. Sci. USA **102**, 13773 (2005).

[27] N. Kashtan, E. Noor, and U. Alon, Proc. Natl. Acad. Sci. USA **104**, 13711 (2007).

[28] W. Kinzel, Phys. Rev. B **33**, 5086 (1986).

[29] B. S. Khatri, T. C. McLeish, and R. P. Sear, Proc. Natl. Acad. Sci. USA **1006**, 9564 (2009).

[30] K. Vetsigian, C. Woese, and N. Goldenfeld, Proc. Natl. Acad. Sci. USA **103**, 10696 (2006).

[31] G. R. Price, Nature **227**, 520 (1970).

[32] L. Peliti, Europhys. Lett. **57**, 745 (2002).

[33] J.-M. Park and M. W. Deem, J. Stat. Phys. **125**, 975 (2006).

[34] T. F. Cooper and R. E. Lenski, BMC Evol. Biol. **10**, 1e11 (2010).

[35] M. Parter, N. Kashtan, and U. Alon, BMC Evol. Biol. **7**, 169 (2007).

[36] A. Kreimer, E. Borenstein, U. Gophna, and E. Ruppin, Proc. Natl. Acad. Sci. USA **105**, 6976 (2008).